Lesson 3.2.2

Least Squares Regression Line as Line of Best Fit

INSTRUCTOR SPECIFIC MATERIAL IS INDENTED AND APPEARS IN GREY

ESTIMATED TIME

1 hour 40 minutes

MATERIALS REQUIRED

Calculator

BRIEF DESCRIPTION

This lesson develops the concepts of the least squares regression line.

LEARNING GOALS

Students will understand that:

- The least squares regression line is the line that minimizes the sum of the squared vertical deviations from the line.

Students will be able to:

- Find the equation of the least squares regression line using technology given a bivariate numerical data set.
- Explain the meaning of *least squares* in a regression setting.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

INTRODUCTION

Students will explore how to identify the line that is the best fit. They will use technology to find the equation of the line and develop and understanding of what it means to say that a particular line is the best fit.

STUDENT MATERIAL IS NOT INDENTED AND APPEARS IN BLACK

INTRODUCTION

Comparing Lines for Predicting Textbook Costs

In the previous lesson, you predicted the value of the response variable knowing the value of the *explanatory variable* (also known as *predictor variable*) using a best-fit line. In the homework you also investigated the concept of extrapolation, which is the idea that, even with the best line, the predictions based on this line may be unreliable if the value of the explanatory variable is outside the range of the data.

So, how do you identify the line that is the best fit? You will use technology to find the equation of the line, but what does it mean to say that a particular line is the best fit? In this lesson, you will investigate this question with the goal of developing a method for determining which line is the best-fit line.

1 Here are the publishers' suggested list prices in 2010 for 12 popular introductory statistics textbooks. The table below gives the descriptive statistics for the price data.

	Min	Q1	Median	Q3	Max	Mean	Standard Deviation
List price	170.95	122.00	150.67	162.55	190.95	147.61	25.72

Stats textbooks

	price
1	150.67
2	122.00
3	149.10
4	166.15
5	107.95
6	181.95
7	158.95
8	151.95
9	122.00
10	150.67
11	190.95
12	118.95

A If someone asks you how much an introductory statistics textbook costs, what prediction would you give? Explain your reasoning.

Answers will vary: Reasonable answers are the mean or median, or perhaps a range mean ± stddev or Q1 or Q3.

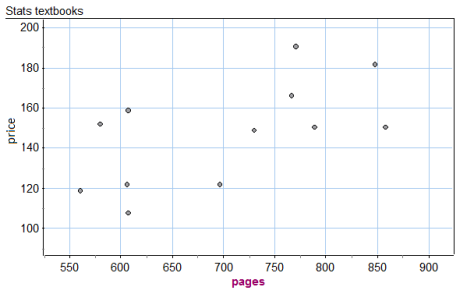
© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

B What variables might be useful for predicting the cost of an introductory statistics textbook?

> **Answers will vary:** Possibilities include both categorical and quantitative variables, such as type of course, hard cover vs. soft cover vs. e-books, number of pages.

C The number of pages in the textbook is one variable you could use to predict price. The scatterplot shows the relationship between pages and price for these 12 textbooks. The data have a somewhat linear form and the correlation coefficient is 0.79, so it makes sense to use a line to summarize the relationship between pages and price. Draw a line that you think is a good summary of the relationship between these two variables. Use the graph of your line to predict the price of a 650-page textbook. Then compare your prediction with a classmate.



2 Since there are infinitely many lines that you could draw, you need a way to determine which line is the best summary of the relationship between two quantitative variables.

You will begin your investigation of how to define a best-fit line by comparing how well four lines predict the list price of the textbooks based on the number of pages.

Stats textbooks

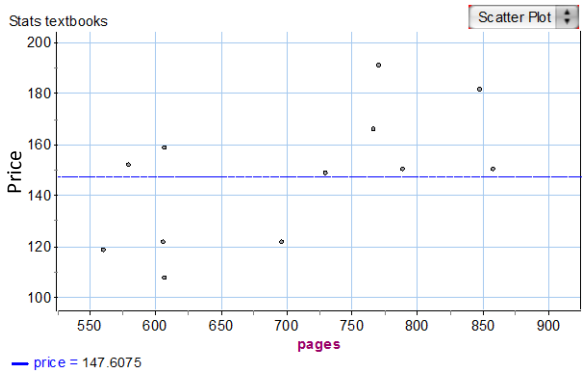
	pages	price	Line_A_predictions	Line_B_predictions	Line_C_predictions	Line_D_predictions
1	560	118.95	147.6075	187.1700	126.3852	109.8900
2	579	151.95	147.6075	182.3915	129.2428	114.6020
3	606	122.00	147.6075	175.6010	133.3036	121.2980
4	607	107.95	147.6075	175.3495	133.4540	121.5460
5	607	158.95	147.6075	175.3495	133.4540	121.5460
6	696	122.00				
7	730	149.10	147.6075	144.4150	151.9532	152.0500
8	766	166.15	147.6075	135.3610	157.3676	160.9780
9	770	190.95	147.6075	134.3550	157.9692	161.9700
10	788	150.67	147.6075	129.8280	160.6764	166.4340
11	847	181.95				
12	857	150.67	147.6075	112.4745	171.0540	183.5460

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

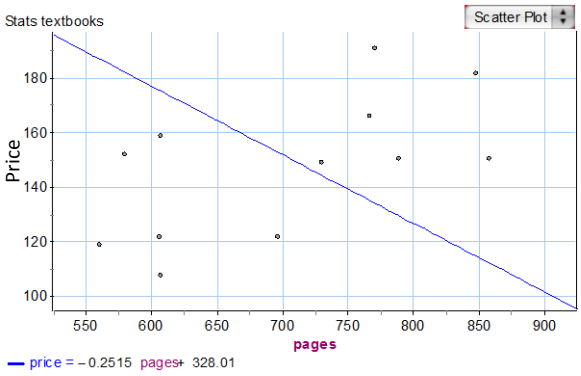
Lesson 3.2.2

Least Squares Regression Line as Line of Best Fit

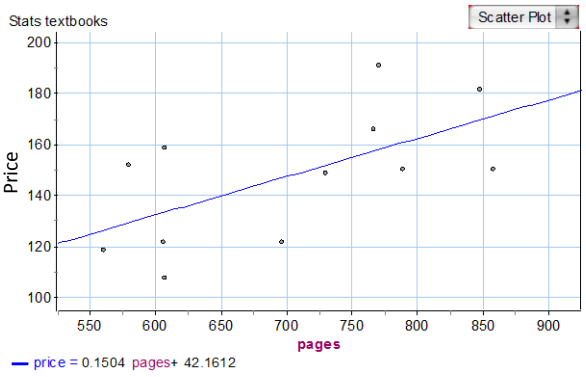
Line A (Mean Price)



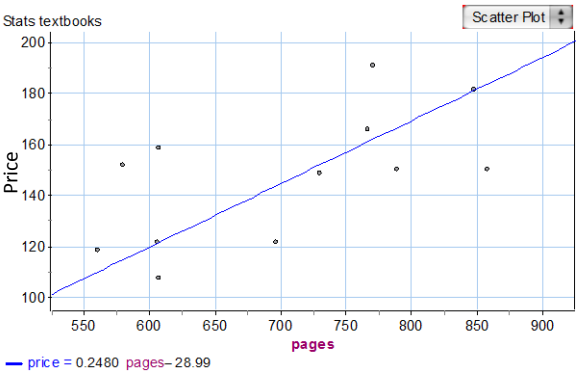
Line B



Line C



Line D



A Begin by using the equation for each line to complete the two incomplete rows in the table of predicted values. (You are predicting prices. It makes sense to write prices with two decimal places, such as $147.61 instead of $147.6075 like you see in the table. You might be wondering why you are recording answers to four decimal places. This is because you will need this level of accuracy to develop some ideas later. So, record your answers to four decimal places for these activities.)

> **Answer:** Line A: 147.6075, 147.6075; Line B: 152.9660, 114.9895; Line C: 146.8396, 169.5500; Line D: 143.6180, 181.0660.

B Which of the four lines do you think results in the best overall predictions of price? Why? How are you selecting the best line?

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

WRAP-UP

Poll the class to see which lines they choose as the best summary. For each line that receives votes as the best summary, call on a few students to explain why they (or their group) chose that particular line. Students may have difficulty articulating their observations as criteria of why one line seems a better summary, so translate their explanations into criteria. Double-check that you have captured the idea they are trying to articulate. Explain how each criterion is visualized in the scatterplot and how it also relates to numerical information in the table. Do not worry if some of their criteria do not characterize regression lines. You can revisit these criteria at the end of the lesson after students have learned about the least squares criterion.

Examples of Criterion from Visual and Numerical Perspectives

	Graph	Table
Possible Criterion	The line should have the same direction as association between the variables. (Slope should be the same sign as the correlation coefficient.)	As the number of pages increases, so does the predicted price.
	The line should come as close as possible to as many points as possible.	The predicted prices are as close as possible to the actual prices. (The differences between predicted price and actual price are as small as possible.)
	The line should go through as many points as possible.	For as many textbooks as possible, the predicted price should equal the actual price.
	The line should have the same number of points above it as below it.	The number of predictions that are greater than the list price equals the number of predictions that are less than the list price.

This table is an example of how to discuss possible criteria. Students might generate other ideas.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

INTRODUCTION

You want a line that minimizes the errors in predictions for individuals in the sample. The reasoning is that if the line is good at predicting the response for the textbooks in the sample, when the response is already known, then it will work well for predicting the response in the future when only the explanatory variable is known.

Now let's get more precise and transform some of the criteria into a numerical measurement that can be used to identify which line gives the best fit.

The idea that you want to develop here is that the line that is the best summary of the relationship between the number of pages and the price of the textbook will give the best predictions for price, but most predictions will have some error. Where possible, tie the criteria generated by the class to the idea of prediction error (e.g., the line gives accurate predictions for points close to it, which means that for a given number of pages, the predicted price is close to the list price).
You can think of the price of the textbook as the price predicted by the linear model plus some error, Price = Prediction + Error. (Some statisticians refer to this idea more generally as Data = Model + Error.) To determine the prediction error, you calculate the difference between price and the predicted price, Price – Prediction = Error.

For the following activities in **Next Steps**, intervene if students need help. Provide guidance as necessary to help students correctly answer the questions about predicted errors.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

NEXT STEPS

Thinking About Prediction Error

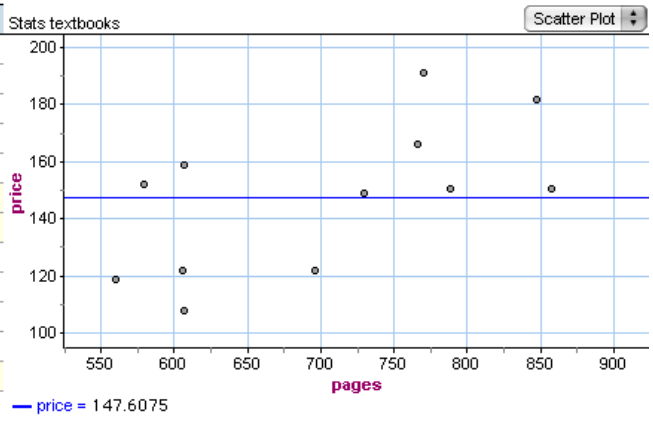
3 For each linear model, complete the missing parts of the table and answer the questions.

A Line A (Mean Price)

> **Answer:** -25.6075, 34.3425.

Stats textbooks_2

	pages	price	Line_A_predictions	Line_A_error
1	560	118.95	147.6075	-28.6575
2	579	151.95	147.6075	4.3425
3	606	122.00	147.6075	-25.6075
4	607	107.95	147.6075	-39.6575
5	607	158.95	147.6075	11.3425
6	696	122.00	147.6075	
7	730	149.10	147.6075	1.4925
8	766	166.15	147.6075	18.5425
9	770	190.95	147.6075	43.3425
10	788	150.67	147.6075	3.0625
11	847	181.95	147.6075	
12	857	150.67	147.6075	3.0625



Identify the following textbooks in the scatterplot and the table:

i The textbook for which the line comes closest to predicting the list price

> **Answer:** Book 7.

ii The textbook for which the prediction is furthest from the list price

> **Answer**: Book 9.

iii In the scatterplot, circle the textbooks that have a negative prediction error. What does a negative error tell you?

> **Answer:** The prediction is higher than the actual value. The prediction line is above the actual value.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

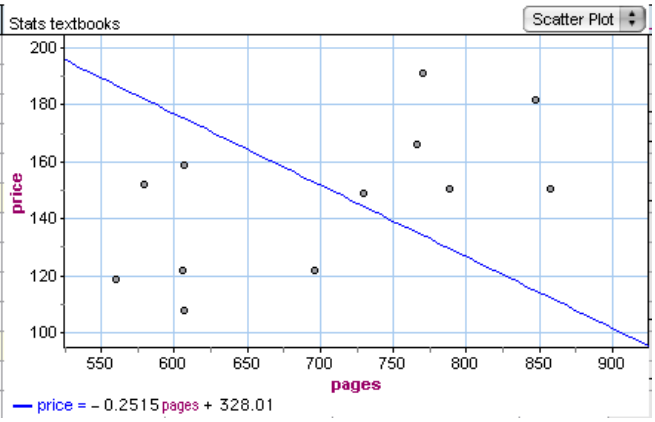
Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

B Line B

> **Answer:** -30.966, 66.9605.

Stats textbooks

	pages	price	Line_B_predictions	Line_B_error
1	560	118.95	187.1700	-68.2200
2	579	151.95	182.3915	-30.4415
3	606	122.00	175.6010	-53.6010
4	607	107.95	175.3495	-67.3995
5	607	158.95	175.3495	-16.3995
6	696	122.00	152.9660	
7	730	149.10	144.4150	4.6850
8	766	166.15	135.3610	30.7890
9	770	190.95	134.3550	56.5950
10	788	150.67	129.8280	20.8420
11	847	181.95	114.9895	
12	857	150.67	112.4745	38.1955



Identify the following textbooks in the scatterplot and the table:

i The textbook for which the line comes closest to predicting the list price

> **Answer**: Book 7.

ii The textbook for which the prediction is furthest from the list price

> **Answer:** Book 1.

iii How can you tell by looking at the scatterplot if the prediction error for a textbook is positive or negative?

> **Answer**: The prediction error is positive when the data value is above the line and negative when the data value is below the line.

iv Identify a textbook for which Line A predicts too low a price but Line B predicts too high a price.

> **Answer:** Book 2 or Book 5.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

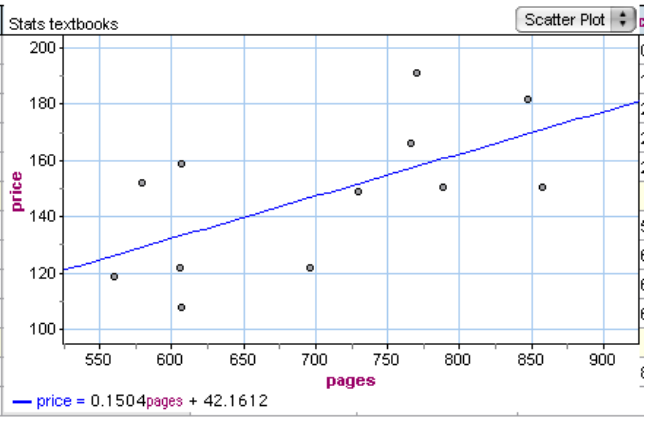
Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

C Line C

> **Answer:** -24.8396, 12.4.

Stats textbooks

	pages	price	Line_C_predictions	Line_C_error
1	560	118.95	126.3852	-7.4352
2	579	151.95	129.2428	22.7072
3	606	122.00	133.3036	-11.3036
4	607	107.95	133.4540	-25.5040
5	607	158.95	133.4540	25.4960
6	696	122.00	146.8396	
7	730	149.10	151.9532	-2.8532
8	766	166.15	157.3676	8.7824
9	770	190.95	157.9692	32.9808
10	788	150.67	160.6764	-10.0064
11	847	181.95	169.5500	
12	857	150.67	171.0540	-20.3840



Identify the following textbooks in the scatterplot and the table:

i The textbook for which the line comes closest to predicting the list price

> **Answer:** Book 7.

ii The textbook for which the prediction is furthest from the list price

> **Answer:** Book 9.

iii All the textbooks for which the predicted list price is within $15 of the actual list price

> **Answer:** Books 1, 3, 7, 8, 10,11.

iv How can you tell by looking at the scatterplot that the prediction error is positive?

> **Answer:** The prediction error is positive when the data value is above the line.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

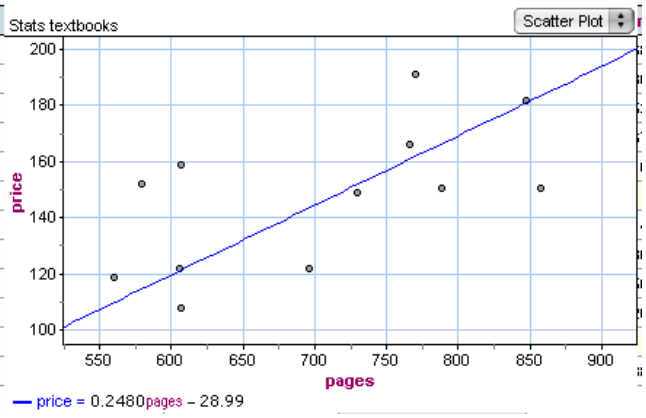
Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

D Line D

> **Answer:** -21.618, 0.884.

Stats textbooks

	pages	price	Line_D_predictions	Line_D_error
1	560	118.95	109.8900	9.0600
2	579	151.95	114.6020	37.3480
3	606	122.00	121.2980	0.7020
4	607	107.95	121.5460	-13.5960
5	607	158.95	121.5460	37.4040
6	696	122.00	143.6180	
7	730	149.10	152.0500	-2.9500
8	766	166.15	160.9780	5.1720
9	770	190.95	161.9700	28.9800
10	788	150.67	166.4340	-15.7640
11	847	181.95	181.0660	
12	857	150.67	183.5460	-32.8760



Identify the following textbooks in the scatterplot and the table:

i The textbook for which the line comes closest to predicting the list price

> **Answer**: Book 3.

ii the textbook for which the prediction is furthest from the list price

> **Answer:** Book 5.

iii all the textbooks for which the predicted list price is exceeds the actual list price by $20 or more

> **Answer**: Books 2, 5, 6, 9, 12.

E The goal is to identify a line that is the best summary of the relationship between pages and price. The best-fit line gives the best predictions of list price, which means that overall it has the least amount of error in the predictions. Rank the four lines from best to worst with the best being the line that gives the best overall predictions of list price. Briefly explain the reasoning behind your rankings.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

WRAP-UP

Frame this part of the lesson as a discussion of how to use the prediction error for each textbook to define a measure of overall prediction error for the line. You will discuss different possible ways to measure overall prediction error, but ultimately you will only use one measure.

Of course, you want to be able to rank the lines from best to worst using the measurement of overall error. Begin by getting a sense of the class's ranking of the lines. If students worked in groups on the previous tasks, you might take a quick tally on the board and then quickly develop a ranking that reflects (as best as you can) the groups' rankings.

Group	1 = best	2	3	4 = worst
1	D	C	A	B
2	C	D	A	B
etc.				

Begin with a simple way to determine the overall error: just sum the errors. Then discuss whether the sum of the errors helps identify the line that best summarizes the data.

When you add up the errors, you get the following sums:

Line	Sum of Errors
A	0
B	–48.8405
C	0.0404
D	32.7460

Does the sum of the errors help identify the line that makes the best predictions? Why or why not? (**Note:** Even if you are conducting this portion of the lesson as a lecture, let students think about this for a minute before you proceed with the following answer to the question.)

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

If you use the sum of the errors to identify the best line, you choose Line A as the best because the cumulative error is zero. Line A uses the mean price as the prediction for price of every textbook. However, this line does not appear to be the best line to summarize the data because the flat mean line does not capture the positive association between pages and price. So, the sum of the errors is a poor way to measure overall error.

Why is the sum of the prediction errors from the mean line zero? You know from your work in Module 2 that the sum of the deviations from the mean is always zero (because positive and negative deviations combine to give a sum of zero). This is why you developed a more sophisticated way to measure spread relative to the mean using variance and standard deviation.

You need to do the same type of thinking here. You need to make the errors all positive to get a sense of the total error. There are two obvious ways to do this:

- Use the absolute value of the errors or
- Square each error and then sum.

Use both strategies to calculate the overall error for the four lines and see which (if either) of the methods helps identify the line that best fits the data.

Have students take notes using the following three tables. Keep students focused on the purpose of this investigation: to develop a measure of overall error that can be used to identify the line that best fits the data. Give students a few minutes to complete the tables.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2

Least Squares Regression Line as Line of Best Fit

	pages	price	Line_A_predictions	Line_A_error	Absolute_value_of_Line_A_error	Line_A_error_squared
1	560	118.95	147.6075	-28.6575	28.6575	
2	579	151.95	147.6075	4.3425	4.3425	
3	606	122.00	147.6075	-25.6075	25.6075	655.7441
4	607	107.95	147.6075	-39.6575		1572.7173
5	607	158.95	147.6075	11.3425		128.6523
6	696	122.00	147.6075	-25.6075	25.6075	655.7441
7	730	149.10	147.6075	1.4925	1.4925	2.2276
8	766	166.15	147.6075	18.5425	18.5425	343.8243
9	770	190.95	147.6075	43.3425	43.3425	1878.5723
10	788	150.67	147.6075	3.0625	3.0625	9.3789
11	847	181.95	147.6075	34.3425	34.3425	1179.4073
12	857	150.67	147.6075			

	pages	price	Line_C_predictions	Line_C_error	Absolute_Values_of_Line_C_error	Line_C_error_squared
1	560	118.95	126.3852	-7.4352	7.4352	55.2822
2	579	151.95	129.2428	22.7072	22.7072	
3	606	122.00	133.3036	-11.3036	11.3036	
4	607	107.95	133.4540	-25.5040	25.5040	650.4540
5	607	158.95	133.4540	25.4960	25.4960	650.0460
6	696	122.00	146.8396	-24.8396		617.0057
7	730	149.10	151.9532	-2.8532	2.8532	8.1408
8	766	166.15	157.3676	8.7824		77.1305
9	770	190.95	157.9692	32.9808	32.9808	1087.7332
10	788	150.67	160.6764	-10.0064	10.0064	100.1280
11	847	181.95	169.5500	12.4000	12.4000	153.7600
12	857	150.67	171.0540			

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

Which measures of the total error help you determine how well a line fits the data?			
Line	*Sum of Error*	*Sum of Absolute Value of Errors (SAE)*	*Sum of Squares of Errors (SSE*
A	0.0000	239.0600	7,275.7566
B	–48.9605	485.0950	24,774.1494
C	0.0404	204.6924	4,458.5762
D	32.7460	206.3540	5,734.1069

Note: Quickly provide answers to the missing parts of the tables with brief explanations as you go.

Use these follow-up questions:

- How does the first row of numbers in the last table relate to the previous table for Line A? **Answer:** These are the sums of the numbers in the previous columns.
- Describe how you calculate the measures of total error given in the last table for Line D. **Answer:** Essentially describe how the values in the columns in the table for Line C are calculated.
- Of the four lines you analyzed, which line is the best summary of the relationship between pages and price if you use the sum of the absolute value of the errors? Which is the second best summary line using this criterion? Third? Fourth? **Answer:** C, D, A, B
- Of the four lines you analyzed, which line is the best summary of the relationship between pages and price if you use the sum of the squares of the errors? Rank the lines from best to worst using this criterion. **Answer:** C, D, A, B

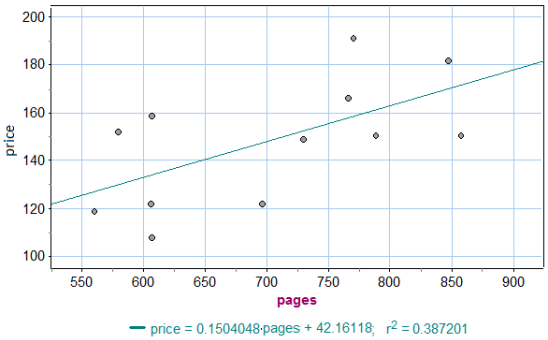
Statisticians square the errors and then find the line that minimizes the sum of the squared errors. The line that has the smallest sum of the squared errors is called the *least squares regression line*. This line minimizes the sum of the squares of the errors, when compared to **all** other possible lines. You will use technology to find the equation of the least squares regression line. In the next lesson, you will learn more about the distinguishing features of this line.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

For these data, Line C is very close to the least squares regression line. Graphically they are indistinguishable. Here is the least squares line shown in the scatterplot.

Now add the least squares line to your table of measurements for total error.



Notice the following when you compare the least squares line to the other lines:

- The sum of the errors is zero for the least squares regression line.
- The sum of the squares of the errors is the smallest for this line (hence the name *least squares*). This is true if you compare the least squares regression line to any other line you created.
- The sum of the absolute value of the errors is not the smallest for the least squares regression line. Line C is a better fit if you use this criterion.

Which measures of the total error help determine how well a line fits the data?			
Line	*Sum of Errors*	*SAE*	*SSE*
A	0.0000	239.0600	7,275.7566
B	–48.9605	485.0950	24,774.1494
C	0.0404	204.6924	4,458.5762
D	32.7460	206.3540	5,734.1069
Least squares regression line	0.0000	204.6985	4,458.5761

Now let's return to the criteria you developed to identify best-fit lines at the beginning of the lesson. Given the discussion today, which of the criteria you developed are not true for the least squares line? Which seem to be valid criteria given your work today?

> After the discussion/lecture, demonstrate how to find least squares regression lines using technology and/or distribute instructions.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2

Least Squares Regression Line as Line of Best Fit

TAKE IT HOME

1 Here you have data collected from students at Los Medanos College in 2009. The variable *units* gives the number of college course units the student reported he or she was taking that semester. The variable *textbooks* gives the amount that the student reported spending on textbooks or other resources required for their courses that semester.

	units	textbooks
1	3	120.25
2	4	65.95
3	9	465.00
4	12	430.00
5	14	396.50
6	16	475.00
7	8	208.00
8	1	5.00
9	6	49.10
10	15	685.00
11	9	220.00
12	4	172.00
13	12	302.00
14	12	460.12
15	12	530.00

A Use technology to find the least squares regression line. (Think carefully about which variable is the explanatory variable.)

> **Answers may vary** slightly based on rounding: textbooks =–42.91 + 38.16 units.

B Use the least squares regression line to predict the amount spent on textbooks for a student taking 12 units.

> **Answers may vary** based on rounding in the regression equation: $415.01.

C Explain why the least squares regression line is considered the line of best fit.

> **Answer:** The least squares regression line minimizes the sum of the squares of the errors..

2 With the following applet, you can draw a line that you think fits the data well and compare your line to the least squares regression line.

www.rossmanchance.com/applets/Reg/index.html

Note: In the applet, errors are called *residuals*. This term comes from thinking about a data point as composed of two parts: the part explained by the regression line (the prediction) and the part that is leftover (called the *residual* or *error*).

A Instructions

1) Check *Your line* and click *Move line*. Follow directions to move the line so that it fits the data well.
2) Check *Show residuals* and record the SAE for your line in the table below.
3) Check *Show squared residuals* and record the SSE for your line in the table.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

4) Check *Regression line*.
5) Check *Show residuals* and record the SAE for the regression line in the table.
6) Check *Show squared residuals* and record the SSE for the regression line in the table.

Line Predicting Height Based on Foot Length	Equation of Line	SAE	SSE
Your line			
Regression line			

> **Answers will vary** for both rows because the applet randomly generates data sets.

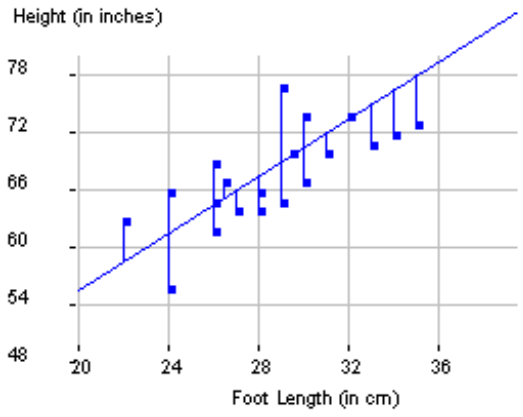
B Compare the values of the SAE and SSE for your line with the regression line. What do you notice? Why does this make sense?

> **Answer:** Students should notice that the SSE, and probably the SAE, is less for the regression line. This makes sense because the regression line is designed to minimize the squared errors. It is the best fit when we use the SSE as a criterion of fit.

C When you click *Show residuals*, you see vertical line segments drawn from each data point to the regression line.

Some line segments are long and others are short. Why is this?

> **Answer:** Some vertical line segments are long because the error is large. The line does not predict a value close to the data value. Similarly, some vertical line segments are short because the error is small.



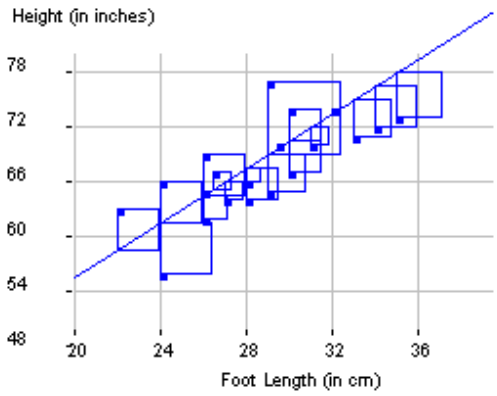
Why do you think these vertical line segments are shown?

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

> **Answer:** These line segments show the size of the error. It is helpful to have a visual way to represent error.

D When you click *Show squared residuals*, you see squares appear.



Some squares are small and others are large. Why is this?

> **Answer:** Some squares are large because the error is large. The squares are drawn based on the length of the vertical line segment representing the error. Similarly, some squares are small because the error is small.

Why do you think these squares are shown?

> **Answer:** These squares show the size of square of the error. It is helpful to have a visual way to represent squared errors.

+++++

This lesson is part of STATWAY™, A Pathway Through College Statistics, which is a product of a Carnegie Networked Improvement Community that seeks to advance student success. Version 1.0, A Pathway Through Statistics, Statway™ was created by the Charles A. Dana Center at the University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. This version 1.5 and all subsequent versions, result from the continuous improvement efforts of the Carnegie Networked Improvement Community. The network brings together community college faculty and staff, designers, researchers and developers. It is an open-resource research and development community that seeks to harvest the wisdom of its diverse participants in systematic and disciplined inquiries to improve developmental mathematics instruction. For more information on the Statway Networked Improvement Community, please visit **carnegiefoundation.org.** For the most recent version of instructional materials, visit **Statway.org/kernel.**

+++++

STATWAY™ and the Carnegie Foundation logo are trademarks of the Carnegie Foundation for the Advancement of Teaching. A Pathway Through College Statistics may be used as provided in the CC BY

Lesson 3.2.2
Least Squares Regression Line as Line of Best Fit

license, but neither the Statway trademark nor the Carnegie Foundation logo may be used without the prior written consent of the Carnegie Foundation.